4 First-Order Logic with Equality

Equality is the most important relation in mathematics and functional programming.

In principle, problems in first-order logic with equality can be handled by any prover for first-order logic without equality:

4.1 Handling Equality Naively

Proposition 4.1 *Let F be a closed first-order formula with equality. Let $\sim \notin \Pi$ be a new predicate symbol. The set $Eq(\Sigma)$ contains the formulas*

$$\begin{gathered}
\forall x\,(x \sim x)\\
\forall x, y\,(x \sim y \rightarrow y \sim x)\\
\forall x, y, z\,(x \sim y \wedge y \sim z \rightarrow x \sim z)\\
\forall \vec{x}, \vec{y}\,(x_1 \sim y_1 \wedge \cdots \wedge x_n \sim y_n \rightarrow f(x_1, \ldots, x_n) \sim f(y_1, \ldots, y_n))\\
\forall \vec{x}, \vec{y}\,(x_1 \sim y_1 \wedge \cdots \wedge x_m \sim y_m \wedge p(x_1, \ldots, x_m) \rightarrow p(y_1, \ldots, y_m))
\end{gathered}$$

for every $f \in \Omega$ and $p \in \Pi$. Let $\tilde{F}$ be the formula that one obtains from F if every occurrence of $\approx$ is replaced by $\sim$. Then F is satisfiable if and only if $Eq(\Sigma) \cup \{\tilde{F}\}$ is satisfiable.

Proof. Let $\Sigma = (\Omega, \Pi)$, let $\Sigma_1 = (\Omega, \Pi \cup \{\sim\})$.

For the "only if" part assume that F is satisfiable and let $\mathcal{A}$ be a Σ-model of F. Then we define a Σ_1-algebra $\mathcal{B}$ in such a way that $\mathcal{B}$ and $\mathcal{A}$ have the same universe, $f_{\mathcal{B}} = f_{\mathcal{A}}$ for every $f \in \Omega$, $p_{\mathcal{B}} = p_{\mathcal{A}}$ for every $p \in \Pi$, and $\sim_{\mathcal{B}}$ is the identity relation on the universe. It is easy to check that $\mathcal{B}$ is a model of both $\tilde{F}$ and of $Eq(\Sigma)$.

The proof of the "if" part consists of two steps.

Assume that the Σ_1-algebra $\mathcal{B} = (U_{\mathcal{B}},\ (f_{\mathcal{B}} : U^n \rightarrow U)_{f \in \Omega},\ (p_{\mathcal{B}} \subseteq U_{\mathcal{B}}^m)_{p \in \Pi \cup \{\sim\}})$ is a model of $Eq(\Sigma) \cup \{\tilde{F}\}$. In the first step, we can show that the interpretation $\sim_{\mathcal{B}}$ of $\sim$ in $\mathcal{B}$ is a congruence relation. We will prove this for the symmetry property, the other properties of congruence relations, that is, reflexivity, transitivity, and congruence with respect to functions and predicates are shown analogously. Let $a, a' \in U_{\mathcal{B}}$ such that $a \sim_{\mathcal{B}} a'$. We have to show that $a' \sim_{\mathcal{B}} a$. Since $\mathcal{B}$ is a model of $Eq(\Sigma)$, $\mathcal{B}(\beta)(\forall x, y\,(x \sim y \rightarrow y \sim x)) = 1$ for every β, hence $\mathcal{B}(\beta[x \mapsto b_1, y \mapsto b_2])(x \sim y \rightarrow y \sim x) = 1$ for every β and every $b_1, b_2 \in U_{\mathcal{B}}$. Set $b_1 = a$ and $b_2 = a'$, then $1 = \mathcal{B}(\beta[x \mapsto a, y \mapsto a'])(x \sim y \rightarrow y \sim x) = (a \sim_{\mathcal{B}} a' \rightarrow a' \sim_{\mathcal{B}} a)$, and since $a \sim_{\mathcal{B}} a'$ holds by assumption, $a' \sim_{\mathcal{B}} a$ must also hold.

In the second step, we will now construct a Σ-algebra $\mathcal{A}$ from $\mathcal{B}$ and the congruence relation $\sim_{\mathcal{B}}$. Let $[a]$ be the congruence class of an element $a \in U_{\mathcal{B}}$ with respect to $\sim_{\mathcal{B}}$. The universe $U_{\mathcal{A}}$ of $\mathcal{A}$ is the set $\{\, [a] \mid a \in U_{\mathcal{B}} \,\}$ of congruence classes of the universe of $\mathcal{B}$. For a function symbol $f \in \Omega$, we define $f_{\mathcal{A}}([a_1], \ldots, [a_n]) = [f_{\mathcal{B}}(a_1, \ldots, a_n)]$, and for a predicate symbol $p \in \Pi$, we define $([a_1], \ldots, [a_n]) \in p_{\mathcal{A}}$ if and only if $(a_1, \ldots, a_n) \in p_{\mathcal{B}}$. Observe that this is well-defined: If we take different representatives of the same congruence class, we get the same result by congruence of $\sim_{\mathcal{B}}$. Now for every Σ-term t and every $\mathcal{B}$-assignment β, $[\mathcal{B}(\beta)(t)] = \mathcal{A}(\gamma)(t)$, where γ is the $\mathcal{A}$-assignment that maps every variable x to $[\beta(x)]$, and analogously for every Σ-formula G, $\mathcal{B}(\beta)(\tilde{G}) = \mathcal{A}(\gamma)(G)$. Both properties can easily shown by structural induction. Consequently, $\mathcal{A}$ is a model of F. $\square$

By giving the equality axioms explicitly, first-order problems with equality can in principle be solved by a standard resolution or tableaux prover.

But this is unfortunately not efficient (mainly due to the transitivity and congruence axioms).

Equality is theoretically difficult: First-order functional programming is Turing-complete.

But: resolution theorem provers cannot even solve equational problems that are intuitively easy.

Consequence: to handle equality efficiently, knowledge must be integrated into the theorem prover.

Roadmap

How to proceed:

- This semester: Equations (unit clauses with equality)

  Term rewrite systems
  Expressing semantic consequence syntactically
  Knuth-Bendix-Completion
  Entailment for equations

- Next semester: Equational clauses

  Combining resolution and KB-completion $\rightarrow$ Superposition
  Entailment for clauses with equality

4.2 Rewrite Systems

Let E be a set of (implicitly universally quantified) equations.

The *rewrite relation* $\rightarrow_E \subseteq \mathrm{T}_\Sigma(X) \times \mathrm{T}_\Sigma(X)$ is defined by

$$\begin{aligned} s \rightarrow_E t \quad \text{iff} \quad & \text{there exist } (l \approx r) \in E,\ p \in \mathrm{pos}(s), \\ & \text{and } \sigma : X \rightarrow \mathrm{T}_\Sigma(X), \\ & \text{such that } s/p = l\sigma \text{ and } t = s[r\sigma]_p. \end{aligned}$$

An instance of the lhs (left-hand side) of an equation is called a *redex* (reducible expression). *Contracting* a redex means replacing it with the corresponding instance of the rhs (right-hand side) of the rule.

An equation $l \approx r$ is also called a *rewrite rule*, if l is not a variable and $\mathrm{var}(l) \supseteq \mathrm{var}(r)$.

Notation: $l \to r$.

A set of rewrite rules is called a *term rewrite system (TRS).*

We say that a set of equations E or a TRS R is terminating, if the rewrite relation $\to_E$ or $\to_R$ has this property.

(Analogously for other properties of abstract reduction systems).

Note: If E is terminating, then it is a TRS.

E-Algebras

Let E be a set of universally quantified equations. A model of E is also called an *E-algebra*.

If $E \models \forall\vec{x}(s \approx t)$, i. e., $\forall\vec{x}(s \approx t)$ is valid in all E-algebras, we write this also as $s \approx_E t$.

Goal:
Use the rewrite relation $\to_E$ to express the semantic consequence relation syntactically:

$s \approx_E t$ if and only if $s \leftrightarrow_E^* t$.

Let E be a set of equations over $\mathrm{T}_\Sigma(X)$. The following inference system allows to derive consequences of E:

$$E \vdash t \approx t \qquad \textit{(Reflexivity)}$$

$$\frac{E \vdash t \approx t'}{E \vdash t' \approx t} \qquad \textit{(Symmetry)}$$

$$\frac{E \vdash t \approx t' \qquad E \vdash t' \approx t''}{E \vdash t \approx t''} \qquad \textit{(Transitivity)}$$

$$\frac{E \vdash t_1 \approx t'_1 \quad \dots \quad E \vdash t_n \approx t'_n}{E \vdash f(t_1, \dots, t_n) \approx f(t'_1, \dots, t'_n)} \qquad \textit{(Congruence)}$$

$$E \vdash t\sigma \approx t'\sigma \qquad \textit{(Instance)}$$
if $(t \approx t') \in E$ and $\sigma : X \to \mathrm{T}_\Sigma(X)$

Lemma 4.2 *The following properties are equivalent:*

(i) $s \leftrightarrow_E^* t$

(ii) $E \vdash s \approx t$ *is derivable.*

Proof. (i)⇒(ii): $s \leftrightarrow_E t$ implies $E \vdash s \approx t$ by induction on the depth of the position where the rewrite rule is applied; then $s \leftrightarrow_E^* t$ implies $E \vdash s \approx t$ by induction on the number of rewrite steps in $s \leftrightarrow_E^* t$.

(ii)⇒(i): By induction on the size (number of symbols) of the derivation for $E \vdash s \approx t$. □

Constructing a *quotient algebra:*

Let X be a set of variables.

For $t \in \mathrm{T}_\Sigma(X)$ let $[t] = \{\, t' \in \mathrm{T}_\Sigma(X) \mid E \vdash t \approx t' \,\}$ be the *congruence class* of t.

Define a Σ-algebra $\mathrm{T}_\Sigma(X)/E$ (abbreviated by $\mathcal{T}$) as follows:

$U_\mathcal{T} = \{\, [t] \mid t \in \mathrm{T}_\Sigma(X) \,\}$.

$f_\mathcal{T}([t_1], \dots, [t_n]) = [f(t_1, \dots, t_n)]$ for $f \in \Omega$.

Lemma 4.3 $f_\mathcal{T}$ *is well-defined: If* $[t_i] = [t'_i]$, *then* $[f(t_1, \dots, t_n)] = [f(t'_1, \dots, t'_n)]$.

Proof. Follows directly from the *Congruence* rule for $\vdash$. □

Lemma 4.4 $\mathcal{T} = \mathrm{T}_\Sigma(X)/E$ *is an* E*-algebra.*

Proof. Let $\forall x_1 \dots x_n (s \approx t)$ be an equation in E; let β be an arbitrary assignment.

We have to show that $\mathcal{T}(\beta)(\forall \vec{x}(s \approx t)) = 1$, or equivalently, that $\mathcal{T}(\gamma)(s) = \mathcal{T}(\gamma)(t)$ for all $\gamma = \beta[\, x_i \mapsto [t_i] \mid 1 \leq i \leq n \,]$ with $[t_i] \in U_\mathcal{T}$.

Let $\sigma = \{x_1 \mapsto t_1, \dots, x_n \mapsto t_n\}$, then $s\sigma \in \mathcal{T}(\gamma)(s)$ and $t\sigma \in \mathcal{T}(\gamma)(t)$.

By the *Instance* rule, $E \vdash s\sigma \approx t\sigma$ is derivable, hence $\mathcal{T}(\gamma)(s) = [s\sigma] = [t\sigma] = \mathcal{T}(\gamma)(t)$. □

Lemma 4.5 *Let X be a countably infinite set of variables; let $s, t \in \mathrm{T}_\Sigma(X)$. If $\mathrm{T}_\Sigma(X)/E \models \forall\vec{x}(s \approx t)$, then $E \vdash s \approx t$ is derivable.*

Proof. Assume that $\mathcal{T} \models \forall\vec{x}(s \approx t)$, i. e., $\mathcal{T}(\beta)(\forall\vec{x}(s \approx t)) = 1$. Consequently, $\mathcal{T}(\gamma)(s) = \mathcal{T}(\gamma)(t)$ for all $\gamma = \beta[\, x_i \mapsto [t_i] \mid 1 \leq i \leq n \,]$ with $[t_i] \in U_\mathcal{T}$.

Choose $t_i = x_i$, then $[s] = \mathcal{T}(\gamma)(s) = \mathcal{T}(\gamma)(t) = [t]$, so $E \vdash s \approx t$ is derivable by definition of $\mathcal{T}$. □

Theorem 4.6 ("Birkhoff's Theorem") *Let X be a countably infinite set of variables, let E be a set of (universally quantified) equations. Then the following properties are equivalent for all $s, t \in \mathrm{T}_\Sigma(X)$:*

(i) $s \leftrightarrow_E^* t$.

(ii) $E \vdash s \approx t$ *is derivable.*

(iii) $s \approx_E t$, *i. e.,* $E \models \forall\vec{x}(s \approx t)$.

(iv) $\mathrm{T}_\Sigma(X)/E \models \forall\vec{x}(s \approx t)$.

Proof. (i)⇔(ii): Lemma 4.2.

(ii)⇒(iii): By induction on the size of the derivation for $E \vdash s \approx t$.

(iii)⇒(iv): Obvious, since $\mathcal{T} = \mathrm{T}_\Sigma(X)/E$ is an E-algebra.

(iv)⇒(ii): Lemma 4.5. □

Universal Algebra

$\mathrm{T}_\Sigma(X)/E = \mathrm{T}_\Sigma(X)/{\approx_E} = \mathrm{T}_\Sigma(X)/{\leftrightarrow_E^*}$ is called the *free E-algebra* with generating set $X/{\approx_E} = \{\, [x] \mid x \in X \,\}$:

Every mapping $\varphi : X/{\approx_E} \to \mathcal{B}$ for some E-algebra $\mathcal{B}$ can be extended to a homomorphism $\hat{\varphi} : \mathrm{T}_\Sigma(X)/E \to \mathcal{B}$.

$\mathrm{T}_\Sigma(\emptyset)/E = \mathrm{T}_\Sigma(\emptyset)/{\approx_E} = \mathrm{T}_\Sigma(\emptyset)/{\leftrightarrow_E^*}$ is called the *initial E-algebra.*

${\approx_E} = \{\, (s, t) \mid E \models s \approx t \,\}$ is called the *equational theory* of E.

${\approx_E^I} = \{\, (s, t) \mid \mathrm{T}_\Sigma(\emptyset)/E \models s \approx t \,\}$ is called the *inductive theory* of E.

Example:

Let $E = \{\forall x(x + 0 \approx x),\ \forall x \forall y(x + s(y) \approx s(x + y))\}$. Then $x + y \approx_E^I y + x$, but $x + y \not\approx_E y + x$.

4.3 Confluence

Let $(A, \to)$ be an abstract reduction system.

b and $c \in A$ are *joinable*, if there is a a such that $b \to^* a \leftarrow^* c$.
Notation: $b \downarrow c$.

The relation $\to$ is called

Church-Rosser, if $b \leftrightarrow^* c$ implies $b \downarrow c$.

confluent, if $b \leftarrow^* a \to^* c$ implies $b \downarrow c$.

locally confluent, if $b \leftarrow a \to c$ implies $b \downarrow c$.

convergent, if it is confluent and terminating.

Theorem 4.7 *The following properties are equivalent:*

(i) $\to$ *has the Church-Rosser property.*

(ii) $\to$ *is confluent.*

Proof. (i)$\Rightarrow$(ii): trivial.

(ii)$\Rightarrow$(i): by induction on the number of peaks in the derivation $b \leftrightarrow^* c$. $\square$

Lemma 4.8 *If* $\to$ *is confluent, then every element has at most one normal form.*

Proof. Suppose that some element $a \in A$ has normal forms b and c, then $b \leftarrow^* a \to^* c$. If $\to$ is confluent, then $b \to^* d \leftarrow^* c$ for some $d \in A$. Since b and c are normal forms, both derivations must be empty, hence $b \to^0 d \leftarrow^0 c$, so b, c, and d must be identical. $\square$

Corollary 4.9 *If* $\to$ *is normalizing and confluent, then every element* b *has a unique normal form.*

Proposition 4.10 *If* $\to$ *is normalizing and confluent, then* $b \leftrightarrow^* c$ *if and only if* $b{\downarrow} = c{\downarrow}$.

Proof. Either using Thm. 4.7 or directly by induction on the length of the derivation of $b \leftrightarrow^* c$. $\square$

Confluence and Local Confluence

Theorem 4.11 ("Newman's Lemma") *If a terminating relation $\rightarrow$ is locally confluent, then it is confluent.*

Proof. Let $\rightarrow$ be a terminating and locally confluent relation. Then $\rightarrow^+$ is a well-founded ordering. Define $P(a) \Leftrightarrow \big(\forall b, c : b \leftarrow^* a \rightarrow^* c \Rightarrow b \downarrow c\big)$.

We prove $P(a)$ for all $a \in A$ by well-founded induction over $\rightarrow^+$:

Case 1: $b \leftarrow^0 a \rightarrow^* c$: trivial.

Case 2: $b \leftarrow^* a \rightarrow^0 c$: trivial.

Case 3: $b \leftarrow^* b' \leftarrow a \rightarrow c' \rightarrow^* c$: use local confluence, then use the induction hypothesis. $\square$

Rewrite Relations

Corollary 4.12 *If E is convergent (i. e., terminating and confluent), then $s \approx_E t$ if and only if $s \leftrightarrow_E^* t$ if and only if $s{\downarrow_E} = t{\downarrow_E}$.*

Corollary 4.13 *If E is finite and convergent, then $\approx_E$ is decidable.*

Reminder:
If E is terminating, then it is confluent if and only if it is locally confluent.

Problems:

Show local confluence of E.

Show termination of E.

Transform E into an equivalent set of equations that is locally confluent and terminating.

4.4 Critical Pairs

Showing local confluence (Sketch):

Problem: If $t_1 \leftarrow_E t_0 \rightarrow_E t_2$, does there exist a term s such that $t_1 \rightarrow_E^* s \leftarrow_E^* t_2$?

If the two rewrite steps happen in different subtrees (disjoint redexes): yes.

If the two rewrite steps happen below each other (overlap at or below a variable position): yes.

If the left-hand sides of the two rules overlap at a non-variable position: needs further investigation.

Question:
Are there rewrite rules $l_1 \rightarrow r_1$ and $l_2 \rightarrow r_2$ such that some subterm l_1/p and l_2 have a common instance $(l_1/p)\sigma_1 = l_2\sigma_2$?

Observation:
If we assume w.o.l.o.g. that the two rewrite rules do not have common variables, then only a single substitution is necessary: $(l_1/p)\sigma = l_2\sigma$.

Further observation:
The mgu of l_1/p and l_2 subsumes all unifiers σ of l_1/p and l_2.

Let $l_i \rightarrow r_i$ $(i = 1, 2)$ be two rewrite rules in a TRS R whose variables have been renamed such that $\mathrm{var}(l_1) \cap \mathrm{var}(l_2) = \emptyset$. (Remember that $\mathrm{var}(l_i) \supseteq \mathrm{var}(r_i)$.)

Let $p \in \mathrm{pos}(l_1)$ be a position such that l_1/p is not a variable and σ is an mgu of l_1/p and l_2.

Then $r_1\sigma \leftarrow l_1\sigma \rightarrow (l_1\sigma)[r_2\sigma]_p$.

$\langle r_1\sigma, (l_1\sigma)[r_2\sigma]_p\rangle$ is called a *critical pair* of R.

The critical pair is *joinable* (or: converges), if $r_1\sigma \downarrow_R (l_1\sigma)[r_2\sigma]_p$.

Theorem 4.14 ("Critical Pair Theorem") *A TRS R is locally confluent if and only if all its critical pairs are joinable.*

Proof. "only if": obvious, since joinability of a critical pair is a special case of local confluence.

"if": Suppose s rewrites to t_1 and t_2 using rewrite rules $l_i \rightarrow r_i \in R$ at positions $p_i \in \mathrm{pos}(s)$, where $i = 1, 2$. Without loss of generality, we can assume that the two rules are variable disjoint, hence $s/p_i = l_i\theta$ and $t_i = s[r_i\theta]_{p_i}$.

We distinguish between two cases: Either p_1 and p_2 are in disjoint subtrees ($p_1 \mid\mid p_2$), or one is a prefix of the other (w.o.l.o.g., $p_1 \leq p_2$).

Case 1: $p_1 \,||\, p_2$.

Then $s = s[l_1\theta]_{p_1}[l_2\theta]_{p_2}$, and therefore $t_1 = s[r_1\theta]_{p_1}[l_2\theta]_{p_2}$ and $t_2 = s[l_1\theta]_{p_1}[r_2\theta]_{p_2}$.

Let $t_0 = s[r_1\theta]_{p_1}[r_2\theta]_{p_2}$. Then clearly $t_1 \to_R t_0$ using $l_2 \to r_2$ and $t_2 \to_R t_0$ using $l_1 \to r_1$.

Case 2: $p_1 \leq p_2$.

Case 2.1: $p_2 = p_1q_1q_2$, where l_1/q_1 is some variable x.

In other words, the second rewrite step takes place at or below a variable in the first rule. Suppose that x occurs m times in l_1 and n times in r_1 (where $m \geq 1$ and $n \geq 0$).

Then $t_1 \to_R^* t_0$ by applying $l_2 \to r_2$ at all positions $p_1q'q_2$, where q' is a position of x in r_1.

Conversely, $t_2 \to_R^* t_0$ by applying $l_2 \to r_2$ at all positions p_1qq_2, where q is a position of x in l_1 different from q_1, and by applying $l_1 \to r_1$ at p_1 with the substitution θ', where $\theta' = \theta[x \mapsto (x\theta)[r_2\theta]_{q_2}]$.

Case 2.2: $p_2 = p_1p$, where p is a non-variable position of l_1.

Then $s/p_2 = l_2\theta$ and $s/p_2 = (s/p_1)/p = (l_1\theta)/p = (l_1/p)\theta$, so θ is a unifier of l_2 and l_1/p.

Let σ be the mgu of l_2 and l_1/p, then $\theta = \tau \circ \sigma$ and $\langle r_1\sigma, (l_1\sigma)[r_2\sigma]_p\rangle$ is a critical pair.

By assumption, it is joinable, so $r_1\sigma \to_R^* v \leftarrow_R^* (l_1\sigma)[r_2\sigma]_p$.

Consequently, $t_1 = s[r_1\theta]_{p_1} = s[r_1\sigma\tau]_{p_1} \to_R^* s[v\tau]_{p_1}$ and $t_2 = s[r_2\theta]_{p_2} = s[(l_1\theta)[r_2\theta]_p]_{p_1} = s[(l_1\sigma\tau)[r_2\sigma\tau]_p]_{p_1} = s[((l_1\sigma)[r_2\sigma]_p)\tau]_{p_1} \to_R^* s[v\tau]_{p_1}$.

This completes the proof of the Critical Pair Theorem. $\square$

Note: Critical pairs between a rule and (a renamed variant of) itself must be considered – except if the overlap is at the root (i. e., $p = \varepsilon$).

Corollary 4.15 *A terminating TRS R is confluent if and only if all its critical pairs are joinable.*

Proof. By Newman's Lemma and the Critical Pair Theorem. $\square$

Corollary 4.16 *For a finite terminating TRS, confluence is decidable.*

Proof. For every pair of rules and every non-variable position in the first rule there is at most one critical pair $\langle u_1, u_2\rangle$.

Reduce every u_i to some normal form u_i'. If $u_1' = u_2'$ for every critical pair, then R is confluent, otherwise there is some non-confluent situation $u_1' \leftarrow_R^* u_1 \leftarrow_R s \to_R u_2 \to_R^* u_2'$. $\square$